THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Checking for asymmetric default dependence in a credit card portfolio: A copula approach

Citation for published version:
Crook, J & Moreira, F 2011, 'Checking for asymmetric default dependence in a credit card portfolio: A copula approach' Journal of Empirical Finance, vol. 18, no. 4, pp. 728-742. DOI: 10.1016/j.jempfin.2011.05.005

Digital Object Identifier (DOI):
10.1016/j.jempfin.2011.05.005

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Empirical Finance

CHECKING FOR ASYMMETRIC DEFAULT DEPENDENCE IN A CREDIT CARD PORTFOLIO: A COPULA APPROACH

Jonathan Crook
Credit Research Centre, University of Edinburgh
29 Buccleuch Place, Edinburgh/UK, EH8 9JS
Telephone: (+44) 0131-650-3802, Fax: (+44) 0131-668-3053
E-mail: J.Crook@ed.ac.uk

Fernando Moreira*
Credit Research Centre, University of Edinburgh
29 Buccleuch Place, Edinburgh/UK, EH8 9JS
Telephone: (+44) 0755-123-7270, Fax: (+44) 0131-668-3053
E-mail: F.F.Moreira@sms.ed.ac.uk

* Corresponding author.

ABSTRACT

Traditional credit risk models adopt the linear correlation as a measure of dependence and assume that credit losses are normally-distributed. However some studies have shown that credit losses are seldom normal and the linear correlation does not give accurate assessment for asymmetric data. Therefore it is possible that many credit models tend to misestimate the probability of joint extreme defaults.

This paper employs Copula Theory to model the dependence across default rates in a credit card portfolio of a large UK bank and to estimate the likelihood of joint high default rates. Ten copula families are used as candidates to represent the dependence structure. The empirical analysis shows that, when compared to traditional models, estimations based on asymmetric copulas usually yield results closer to the ratio of simultaneous extreme losses observed in the credit card portfolio.

Copulas have been applied to evaluate the dependence among corporate debts but this research is the first paper to give evidence of the outperformance of copula estimations in portfolios of consumer loans. Moreover we test some families of copulas that are not typically considered in credit risk studies and find out that three of them are suitable for representing dependence across credit card defaults.

JEL classification: G20, G21, C14, C46

Keywords: Credit risk, asymmetric dependence, consumer loans, copulas

1. INTRODUCTION

Many credit risk models assume that returns from loans are normally distributed, not only individually but also at the portfolio level. This implies relatively fewer occurrences of simultaneous extreme values than if more appropriated distributions were used and therefore may lead to biased estimations if returns do not follow that particular distribution.

Since the 1960's there is abundant evidence in the literature showing that asset returns in general are not normally distributed (see Mandelbrot,1963 and Fama,1965). Since then many empirical studies have confirmed this behavior for several classes of investments, including loan portfolios (Rosenberg and Schuermann,2006). Moreover, it has also been found that returns are more correlated in the left tail (i.e. when investments result in losses or lower returns) than in the right tail. See, for instance, Ang and Bekaert (2002), Ning (2010) (who cites many other studies that reach the same conclusion) and Patton (2006). According to Di Clemente and Romano (2004) and Das and Geng (2006), returns of credit assets also present asymmetric (tail) dependence.

Copulas are an effective way of capturing diverse dependence structures regardless of the individual distributions and symmetry. They have been used in finance since the end of the 90's and started being applied to credit risk a couple of years later. However in this latter field the application of copulas has been concentrated on corporate debts and derivatives.

The first contribution of this paper is the empirical estimation of best-fit copulas for consumer loans by using a credit card dataset provided by a large UK bank. Then, estimations of joint extreme default rates based on copulas are compared to estimations conditional on the assumption of normality. The second contribution is the test of five copulas that are not usually included in research pertaining to credit risk.

A third innovation is the use of goodness-of-fit tests (GoF) based only on the *right tail* of the variables' distributions (instead of the usual procedures that consider whole distributions). This strategy was implemented because the principal objective of finding the best-fit copulas here is to employ them to estimate the probability of simultaneous high defaults.

A sample of credit card loans was split into five segments according to a score provided by the Bank. Then the association between each of the ten pairs of segments was modeled by the best-fit copula. Most of the pairs of segments present right-tail dependence which suggests the existence of flaws in estimations of joint high defaults derived from traditional models. In other words, such structure means that higher default rates are more associated and the Bank is subject to larger losses in downturns than would be calculated with traditional techniques. We also find that some of the pairs have dependence appropriately represented by three of the five less popular copulas inserted in this study.

After finding the best representation for the dependence across the credit card loans, we compare estimations of conjunct high default rates following conventional assumptions of multivariate normality and Copula Theory. In most cases, the latter method generates values closer to the observed default rates in the dataset[1]. Considering each pair of segments separately and six risk levels (loss percentiles), the copula approach gave overall better results for all pairs.

The remainder of this paper is organized as follows. Section 2 contains a brief review of copulas, their application in credit risk and techniques to estimate copula parameters and decide which copula is the best one among many candidates. Next, we describe the data used in the empirical analysis. The ten copula families taken as candidates to represent the dependence structure among credit card loans are introduced in Section 4. Then, we estimate the dependence structure (copulas) between pairs of segments in a credit card portfolio of a large UK bank. Section 6 compares estimations of joint high defaults in the portfolio studied according to two approaches: by assuming normality and by using the best-fit copula. Final comments are in the last section.

2. COPULAS, TAIL DEPENDENCE AND CREDIT RISK

2.1 Copulas and tail dependence

Copulas are functions that link univariate distributions to the multivariate distribution of the related variables. Let *x* and *y* be random variables. Then a copula may be represented as:

$$H(x,y) = C(F_X(x), G_Y(y))$$

where $F_X(x)$ and $G_Y(y)$ are the cumulative distribution functions of X and Y evaluated at x and y respectively and $C(.)$ is the copula that links those distributions in order to form the joint distribution $H(x,y)$. So, the copula C gives the probability that X and Y are simultaneously below x and y, i.e. $\Pr(X < x, Y < y)$, regardless of the shape of the distributions F_X and G_Y.

[1] Albeit the difference between the estimations via the two approaches was usually not statistically significant (see Section 6).

The probability that the variables are *above* specific points may be found by the survival copula, represented by $\hat{C}$:

$$\Pr(X > x, Y > y) = 1 - F_X(x) - G_Y(y) + \Pr(X < x, Y < y) = \hat{C}(\overline{F}_X(x), \overline{G}_Y(y))$$

where $\overline{F}_X(x) = 1 - F_X(x)$ and $\overline{G}_Y(y) = 1 - G_Y(y)$. Introductory explanations on Copula Theory may be found, for instance, in Joe (1997) and Nelsen (2006).

Copulas allow us to identify different levels of dependence across the distribution (i.e. when different levels of the variables present diverse association). In this paper, we are interested in the upper (right) tails of the default rate distributions and in the dependence present in this particular region (i.e. association among "high" default rates). This relationship can be measured by means of the upper tail dependence parameter, λ_U, given by (see, e.g., Joe, 1997 and Trivedi and Zimmer, 2007):

$$\lambda_U = \lim_{p \to 1^-} \Pr[X > F_X^{-1}(p) \mid Y > G_Y^{-1}(p)] = \lim_{p \to 1^-} \Pr[Y > G_Y^{-1}(p) \mid X > F_X^{-1}(p)]$$

where p is the extreme percentile considered and F_X^{-1} and G_Y^{-1} are the inverse distributions of X and Y respectively. So, λ_U is the probability of one of the variables (X, for example) being greater than a specific percentile in its marginal distribution (F_X) given that the other variable (Y) is greater than that same percentile in its individual distribution (G_Y). Whenever $\lambda_U \in (0,1]$, the variables present upper tail dependence and there is no upper tail dependence if $\lambda_U = 0$. The lower tail dependence parameter, λ_L, can be similarly calculated for variables smaller than specific cutoffs when the percentile p approaches zero. When $\lambda_L > 0$, the data is lower-tail dependent. Both tail dependence parameters are directly associated to the parameters of some copula families (see, for instance, Nelsen, 2006 and Nikoloulopoulos et al., 2011).

2.2 Application of copulas to credit risk

The concept of copula functions was first published in 1959 and was first applied in Finance at the end of the 90's. To the best of our knowledge, Li (2000) was the first paper to apply copulas to study default correlation between two credit risks albeit the author concentrated on derivative products.

After Li (2000), the literature dealing with the use of copulas in credit risk has been focused on corporate debts (Hamilton et al., 2001, Hamerle and Rösch, 2005, and Das and Geng, 2006), derivatives (Melchiori, 2003, Cherubini et al., 2004, and Hull and White, 2006), general theoretical models (Schönbucher and Schubert, 2001 and Kostadinov, 2005) and the relationship between risk factors and defaults (Frey and McNeil, 2001, Frey et al., 2001, Daul et al., 2003, and Schmidt, 2003). The most frequently considered copulas have been the elliptical (mainly Gaussian and Student t) and the Archimedean (especially Clayton, Frank, and Gumbel). Hence, copulas have not been applied in the context of *consumer* loans and other families of copulas (besides the five ones just cited) could also be tested in credit risk analyses.

2.3 Finding the best-fit copula

2.3.1 Parameter estimation techniques

Essentially, there are three parametric approaches to estimate copulas from data: the Exact Maximum Likelihood (ML) method[2], the Inference Functions for Margins (IFM), and the Canonical Maximum Likelihood (CML) method.

In short, ML involves maximizing a function that includes parameters for both the marginal distributions and the copula. The IFM method maximizes two log-likelihood functions. First, the parameters of the margins are found then these values are used to find the copula parameters. The CML also has two stages but the dataset (default rates in this case) is converted into uniform variables, so that it is not necessary to estimate the margins' parameters. Then, in a second step, the copula parameters are estimated by maximizing a log-likelihood function that includes the uniform variables and the copula parameters[3].

Although Kole et al. (2007) state that there is no consensus on the best way to fit copulas to data, Durrleman et al. (2000) found the CML to be the best method to model both simulated and real financial data whilst ML and IFM estimations were biased. Genest et al. (2009) state that the IFM approach is less efficient because it is subject to flawed estimations of the univariate

[2] Also known as Full Maximum Likelihood (FML).

[3] For more details on these three methods, see, for example, Cherubini et al. (2004), McNeil et al. (2005), Genest and Favre (2007), and Trivedi and Zimmer (2007).

distributions (margins) which compromises the second step, namely the search for the copula parameter. Furthermore Cherubini et al. (2004) point out that the ML method tends to be very computationally intensive given that several parameters (for individual distributions and the copula) must be estimated at once. Therefore we will employ the CML method to estimate the parameters of the candidate copulas to represent the dependence across credit card default rates.

2.3.2 Model selection

After finding the parameter of each copula, it is necessary to decide which family is the best representation for the data dependence. There are a few techniques to select the best copula. One of them is based on distance measures pertaining to candidate models' (copulas') distributions and the empirical data's distribution (as implemented by Kole et al., 2007). Other alternatives mentioned, for example, in Patton (2009) are Likelihood Ratio tests[4] and approaches related to information criteria, such as Akaike and Schwarz's Bayesian Information Criteria. Conducting estimations based on more than one of these methods would be interesting to increase the robustness of our results but, since this study does not have the purpose of comparing the power of model selectors with *different structures*, we left this as a future exercise and adopted only one class of those methods. To explore more options of a particular type of model selectors, we chose the class that has more distinct approaches, i.e. the techniques founded on goodness-of-fit (GoF) tests that evaluate the closeness between each candidate copula's distributions and the observed data's distribution.

According to simulations run by Genest et al. (2009) and Berg (2009), the three GoF methods that presented the highest performances were based on: Empirical Copula (the best), Kendall's Transform and Rosenblatt's Transform, which are the three approaches tested by Weiß (2009)[5]. Although we report the results based on those three GoF methods, only the results founded on the Empirical Copula are discussed in depth. Thus, for each pair of segments analyzed, three copulas are designated as the best potential representation of its dependence structure (i.e. each copula identified in line with one of the three GoF tests used: Empirical Copula, Kendall's Transform and Rosenblatt's Transform). It is possible (and expected) that the three approaches

[4] In our case, the most appropriate Likelihood Ratio test would be a pseudo-likelihood test for *non-nested* hypotheses because we used percentiles of the variables (so, a pseudo-likelihood test) and none of the copula families studied is a special case of the other copulas (so, the hypotheses are non-nested).

[5] Readers should consult these three papers for details on the mentioned GoF methods.

yield conflicting results in some cases (i.e. two or three distinct copulas for the same pair). In these circumstances, the Empirical-Copula-based test must be considered the most reliable since, according to Genest et al. (2009), it is the method that presents the least data transformation and its superiority was confirmed by Berg (2009).

In order to verify the significance of the GoF tests, Genest et al. (2009) and Berg (2009) present some routines to calculate p-values concerning the null hypothesis that the dataset dependence (copula) is equal to the tested copula. The procedure to find the p-values consists of simulating the candidate copulas many times with their respective parameters found and checking which proportion of the simulated copulas is "farther" from the empirical data than the candidate copula with the exact parameter found via maximization. Thus high p-values suggest that the considered copula cannot be rejected because it is closer to the observed dataset than most of the other simulated copulas.

3. DATA DESCRIPTION

This empirical study is based on a random sample from the credit card portfolio of a large UK bank comprising the monthly payment status of 177,234 accounts[6] over the period April/2007 – March/2009.

The dataset was split into five segments according to the loans' credit quality (credit score provided by the Bank) in the first month. Each segment corresponds to a quintile of the score distribution such that the least risky segment (named "A") presents the highest scores and the risk level increases with the reduction of the scores up to the riskiest segment (called "E"). Since we used only credit scores of the first month, the loans remain in the same segment until the last month (i.e. there is no migration across the segments over time)[7]. Obviously, our findings concerning the representation of the dependence among the segments (Section 5) are specific for the number of segments analyzed (five) and will reflect any inaccuracy in the scoring process.

Default was defined as the non-payment of three monthly installments (consecutively or not) conditional on no prior default. Thus the default rate for each segment in a particular month was

[6] The dataset initially comprised 350,066 credit card loans but the credit score for 172,832 accounts was not available. From these accounts, 95,052 were not open in the first month covered by the dataset.

[7] According to our definition of segments (score quintiles), if we opted to use the credit scores of every month to update the segments' composition, some distortions would possibly occur. For instance, some loans with equal scores could be set in different segments in different months or some loans with intensely decreasing (or increasing) scores could remain in the same segment over the whole period.

calculated as the amount of loans that reached their third month in arrears for the first time (i.e. conditional on no prior default) divided by the number of active accounts in that month. Once the loans defaulted they were excluded from the dataset (i.e. default is considered an absorbing state).

This procedure generated a time series of default rates with 24 observations for each segment and these values were used to estimate the dependence between the segments. Figure 1 show the joint default rates in all pairs of the credit segments considered (each dot in the scatter plots corresponds to a month for which the default rates are indicated on the axes). The summary statistics of the data are given in Table 1.

[Insert Figure 1 here]

[Insert Table 1 here]

As expected, the mean default rates increase with decreases in the credit quality (from segment A to E). The data dispersion (measured by the standard deviation) has similar behavior.

According to the two rightmost columns, the three first segments are closer to the normal distribution than the two riskiest ones. This is confirmed by the Jarque-Bera test (see Jarque and Bera, 1987) which tests the null hypothesis that a sample comes from a normal distribution (see Table 2 where higher values of the Jarque-Bera statistics lead to the rejection of the null hypothesis).

[Insert Table 2 here]

Although three of the segments may be satisfactorily represented by the normal distribution, this does not imply that the dependence between pairs involving two of those segments will be better expressed by the Gaussian (Normal) copula. As we will show later, even data with normally distributed margins may have diverse dependence structure.

For simplicity, we assume that the default rates of each segment do not exhibit serial correlation (dependence) and the shape of their marginal distributions does not change over time.

4. FAMILIES OF CANDIDATE COPULAS

Ten copula families are tested to represent the dependence across default rates. They were selected from the nine bivariate one-parameter families depicted in Joe (1997, Chapter 5) that are absolutely continuous[8] along with the Student t copula.

Table 3 lists them and their main features in terms of structure. The cumulative distribution functions (*cdfs*) of the candidate copulas can be seen in Joe (1997) and Nelsen (2006).

[Insert Table 3 here]

In sum, the four first copulas in Table 3 indicate that the variables (default rates, in this study) have the same level of dependence below and above their mean and there is no higher association (when compared to the multivariate normal distribution) among extreme values. The Student t is also symmetric but points out more intense relationship among extreme events. The Clayton copula indicates that smaller values are more linked and the other four copulas express the opposite: higher values are more associated. This last case is the one that brings more concern with respect to a credit portfolio since it means that higher default rates tend to happen together more often, i.e. higher losses in each segment occur at the same time and the lender is more subject to financial deficits. All the copulas that present right tail dependence (exclusively or in conjunct with left tail dependence) are associated to an upper tail dependence parameter λ_U (explained in Section 2.1) greater than zero.

5. DEPENDENCE STRUCTURE IN THE CREDIT CARD PORTFOLIO

5.1 Estimation based on the complete default rate distribution

To find the best-fit copula for each of the ten pairs of segments, the parameters of the candidate copulas were estimated for each pair according to the Canonical Maximum Likelihood method which, according to the literature, outperforms the other approaches mentioned in Section 2.3.1. Then the three goodness-of-fit (GoF) tests cited in Section 2.3.2 were used to define which copula better represents the dependence in each pair.

[8] This property is desirable in order to simulate such copulas (used in the goodness-of-fit tests) and estimate their parameters. Both procedures demand derivations of the copulas' cumulative distribution functions.

Table 4 displays the best copulas in the upper-right triangle along with the linear (Pearson's) and a rank (Kendall's tau, in parenthesis) correlation in the lower-left triangle. The copulas displayed for each pair were chosen following goodness-of-fit tests based on the Empirical Copula Method (first family shown), the Kendall's Transform (in parenthesis), and the Rosenblatt's Transform (in square brackets). The parameters of the best-fit copulas are shown in Appendix A.

[Insert Table 4 here]

The *rejection* level of the estimations is indicated by ** and * which represent the levels 5% and 10%, respectively (for instance, although the hypothesis of Clayton copula for the pair AB based on the Empirical GoF approach is the best among the ten alternatives, it can be rejected at the 5% level, i.e. with 95% of confidence). For the sake of brevity, we omit the detailed outcome of all three GoF approaches and their respective p-values. This information is available upon request. The analysis of the results will be based on the copulas estimated following the Empirical Copula method given that it was found in the literature to be the most robust amongst the three models considered in this study (see, as mentioned in Section 2.3.2, Berg, 2009 and Genest et al., 2009).

Eight pairs are represented by copulas that denote tail dependence (which implies $\lambda_U > 0$ and/or $\lambda_L > 0$ as defined in Section 2.1) being that in two cases the estimations can be rejected at the 5% and 10% levels (pairs AB and AC, respectively). This means that in most of the cases the link across extreme default rates (lower and/or higher) is stronger than assumed by the Gaussian copula (which is implicit in traditional credit risk models).

Three of the pairs (AD, BD, and CD) exhibit right-tail dependence, meaning that higher default rates are more associated than the other levels which may strengthen the Bank's losses in downturns. Three other pairs (AB, AC, and BC) have more intense relationships among low default rates, expressing the most profitable scenario for the Bank inasmuch as most of its debtors tend to keep up with their repayments simultaneously in upturns whilst delinquencies are not very related in downturns.

Pairs AE and BE may present those two effects on the Bank's results. The symmetric tail association represented by the Student t implies that both lower and higher ranks of defaults are more associated than intermediate rates.

The dependence in the two riskiest pairs of segments (CE and DE) is better represented by copulas that do not express tail dependence (i.e., $\lambda_L = \lambda_U = 0$)[9]. This condition is beneficial for the lender because the highest default levels do not get more linked in downturns. In fact, these results are counterintuitive given that we expected a higher connection across these segments in extreme situations and, therefore, the best-fit copulas found in our analysis could indicate potential weaknesses in the Bank's score model. However the mean default rates presented in Table 1 corroborate the Bank's evaluation. Since we do not have access to the method used by the Bank to classify the loans we are not able to explain this supposed contradiction.

Apart from the pairs involving the segment E, all dependence structures are better represented by asymmetric copulas (although many of the symmetric families were not rejected in the GoF tests). This suggests that the association of that riskiest segment with the other loans tends to have similar intensity in opposite economic scenarios (booms or crashes). Thus the same level of connections among the riskiest debtors and the other loans in downturns (that raise the losses) may be expected in upturns (so that losses are reduced and profits are potentially amplified).

It is interesting to note that even when the individual distributions (the default distribution for each segment in this study) are satisfactorily approximated by the normal distribution, their joint distribution may not be expressed by the Gaussian copula. This is the case of the pairs AB, AC, and BC (although the estimation of the first two can be rejected).

The p-values for the goodness-of-fit tests (not displayed) entail the non-rejection of many copula families which is possibly a consequence of the short dataset used (24 observations) that is not enough to form unambiguous patterns that match unique joint distributions for each pair of segments. Thus the main inference from this empirical analysis is not the rejection of the Gaussian dependence but the identification of other families that may represent the dependence among credit card loans more accurately and improve the estimations of joint extreme defaults. The dependence in nine of the pairs is better expressed by other copulas; only the pair CE has the Gaussian as the best-fit copula.

5.2 Estimation based on the upper tail of the default rate distribution

Since the main purpose of estimating the dependence structure is to calculate joint "high" defaults, it is possible that a better performance of the copula approach may be achieved if the

[9] Bear in mind that the copulas that indicate tail dependence were not rejected for these pairs.

best-fit copulas are found considering only the right tail of the default rate distributions. In this section, we estimate the copula whose right tail (here, defined as above the 75th percentile of each marginal variable, i.e. the default rates in each segment) gives the best fit to the right tail of the empirical distribution. In this bivariate case, the Empirical Copula $\hat{C}_R$ limited to "high" percentiles (above 0.75 in our example) is calculated from the dataset as:

$$\hat{C}_R(\mathbf{u} \mid \mathbf{u} \geq 0.75) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbf{1}(U_{i1} \leq u_1, U_{i2} \leq u_2)$$

where the term $\mathbf{u} \geq 0.75$ indicates that both marginals u_1 and u_2 must be equal to or greater than 0.75. If $\mathbf{R}_{ij}$ is the i^{th} rank of a random variable X_j with n observations ($1 \leq i \leq n$), $\mathbf{U}_{ij} = \mathbf{R}_{ij} / (n+1)$ will be a pseudo-observation equivalent to that rank normalized to (0,1) where the scaling factor (denominator n+1) is employed to guarantee $\mathbf{U}_{ij}$ in (0,1). $\mathbf{1}$ is an indicator function that returns 1 if all conditions in parenthesis are satisfied and 0 otherwise. $\hat{C}_R$ is, in fact, the conventional Empirical Copula (as described, for instance, in Nelsen, 2006, p. 219) applied only to percentiles above a specified threshold (0.75 in this study). We recognize that this percentile is too low for typical definitions of "high" events but it was chosen because we needed to specify a certain number of extreme occurrences in each segment that allowed us to observe joint events across high defaults and even though we chose only the six highest values of each default distribution as its extreme occurrences, this number represents 25% of each segment's distribution.

Each candidate copula, $C_{R\hat{\theta}}$, pertaining to this test applied to the rightmost region of the default distributions will be evaluated for the same percentiles considered in the Empirical Copula $\hat{C}_R$ and will be given by[10]:

$$C_{R\hat{\theta}} = C_{\hat{\theta}}(\mathbf{u} \mid \mathbf{u} \geq 0.75)$$

[10] Note that the "tail" of the distributions we are modeling is formed not only by the concurrent events $u_1 \geq 0.75$ and $u_2 \geq 0.75$ but also by $u_1 \geq 0.75$ and $u_2 < 0.75$ (and vice versa).

where $C_{\hat{\theta}}$ is a candidate copula with a parameter $\hat{\theta}$ estimated according to the procedures described in Section 2.3.1 and limited, in this example, to percentiles equal to or higher than 0.75.

Then we use $\hat{C}_R$ and $C_{R\hat{\theta}}$ to calculate the Cramér-von Mises statistic, $\hat{E}_R$ (as done for the conventional goodness-of-fit test based on the Empirical Copula for complete distributions – see Genest et al., 2009 and Berg, 2009):

$$\hat{E}_R = n\int_{[0,1]^d} \{\hat{C}_R(\mathbf{u}) - C_{R\hat{\theta}}(\mathbf{u})\}^2 d\hat{C}_R(\mathbf{u}) = \sum_{i=1}^{n} \{\hat{C}_R(\mathbf{u}_i) - C_{R\hat{\theta}}(\mathbf{u}_i)\}^2$$

where $\mathbf{u} \geq 0.75$. The smaller $\hat{E}_R$ is, the more representative (of the empirical data) the respective copula is. So, this test will point out the copula family that has the area in the right tail (when the percentiles u_1 and/or u_2 are greater than or equal to 0.75) closest to the area in the right tail of the joint distribution of the observed data.

The best-fit copulas selected in accordance with GoF tests based on complete default distributions (as in Section 5.1) are supposed to be the best representation of joint distributions in general (i.e. for all values of u_1 and u_2) and may not be the best approximation of the upper tail specifically (which can be given by another copula family). On the other hand, copulas chosen according to $\hat{E}_R$ are the best approximation of "high" values of default rates and might not be the best representation of default rates smaller than the respective 75th percentiles. This approach seems to be an original way to estimate copulas to express joint high events since such strategy has not been found in the literature.

The best-fit copulas based on this alternative method are displayed in Table 5 and their respective parameters are in Appendix B.

[Insert Table 5 here]

Similar to the previous table, the dependence measures in the lower-left triangle are the linear correlation (above) and the Kendall's tau (in parenthesis). The copulas for each pair of segments are assessed from the Empirical Method (first family displayed for each pair), Kendall's

Transform (in parenthesis), and Rosenblatt's Transform (in square brackets). The results for the three GoF approaches and the related p-values are not presented but are available on request.

The results based on the Empirical Copula Method[11] reveal that seven out of the ten pairs present tail dependence, i.e. the association between default rates in extreme cases is more intense than assumed by the Gaussian copula (which suggests that the tail dependence parameters defined in Section 2.1, λ_U and λ_L, are greater than zero in those seven cases). Six of the pairs have right-tail dependence and only one pair, CD, has low default rates more related than the other levels of default rates. Therefore these estimations indicate that most of the associations among credit card loans lead to accentuated losses in adverse scenarios.

The pairs that did not present tail dependence were exactly the ones with negative dependence (all of them involving the riskiest segment, E[12], which is advantageous for the Bank since its highest expected losses are mitigated by better performance of other segments). Note that even in these three instances the best-fit copula was not the Gaussian one. This has some implications only in the central region of the default distribution and does not impact investigations concentrated in extreme events (which is the case of this study).

Compared to the estimations in the prior section (founded on complete default distributions), the results derived from the fit of the right tail allowed a greater rejection of the Gaussian copula (especially with reference to the Kendall's Transform GoF approach according to which that copula can be rejected in two pairs when using the complete default distributions and in six pairs when only the right tails of those distributions are considered).

However, it is important to note that, in accordance with the Empirical Copula method, the copulas tested were typically not rejected at the 5% significance level (the only exception was the rejection of the Galambos copula for the pair BC). We checked the consistency of this lack of rejections by running Likelihood Ratio tests as proposed by Kupiec (1995) which focus on the number of extreme events to determine the suitability of models (copula families in our case). In each of these tests, we tested the null hypothesis that the considered copula family returns the same number of extreme events (*joint* occurrences of default rates above the 75th percentile of their respective marginal distributions) as the number of extreme events present in the dataset. The Likelihood Ratio tests corroborated the results obtained from the Empirical Copula GoF as

[11] As before, the estimations are interpreted with respect to this approach due to its superior robustness according to the pertinent literature.

[12] The negative dependence regarding the segment E can be visually confirmed in Figure 1.

the only rejection was related to the Galambos copula in the pair BC[13]. This impossibility of distinguishing among the copula families is likely due to the small number of observations in our sample.

Another way to identify the copulas that provide the best representation of the relationship across high default rates is to compare the tail dependence parameter λ_U (explained in Section 2.1) of the empirical dataset to the λ_U implied by each of the candidate copulas. For five of the families included in this study (Gaussian, Frank, FGM, Placket, and Clayton), $\lambda_U = 0$ (since these families do not have upper tail dependence; see Nelsen, 2006) and for another three (Gumbel, Joe and Student t) λ_U can be directly estimated from the copula's parameter θ (see Nelsen, 2006, p. 215, for the first two copulas and Nikoloulopoulos et al., 2011, for the Student t).

Thus, this additional test has fewer options (candidate copulas) than the other two prior techniques (GoF and Likelihood Ratio test) since there is no closed-form expression to derive λ_U from the parameters θ of two of the candidate families that express right tail dependence (Galambos and Hüsler-Reiss). Furthermore the concept of λ_U implies the use of percentiles close to the 100th percentile for the right tail dependence but our dataset has only 24 observations and the farthest point we can check is the 96th percentile. Given these limitations, our results based on this approach should be seen with reservation but even though we present them here to show this alternative way to empirically choose copulas when we have special interest in one or both tails.

First of all, we could not distinguish among the five families that do not entail upper extreme dependence (Gaussian, Frank, FGM, Plackett, and Clayton) as all of them have $\lambda_U = 0$. This was the case of pairs AD, AE, BE, and CE besides the fact that the Gumbel copula was also a good approximation for these four pairs since its λ_U was very close to zero (1.88 x 10^{-6}).

Conversely, the best representation for the other six pairs indicated right tail dependence: Gumbel (pairs AB, AC and CD), Joe (BC and BD) and Student t (DE). So, in spite of the limitations of this λ_U-based approach, it supports our earlier conclusion according to which most of the pairs of segments present upper tail dependence (i.e. higher defaults rates are more associated than other levels of default rates).

---

[13] Regarding the 10% significance level, other copulas were rejected in accordance with both the Empirical Copula GoF (Student t, Gumbel and Hüsler-Reiss in the pair BC and Student t in the pair DE) and the Likelihood Ratio tests (Clayton, Hüsler-Reiss and Galambos in the pair BE and Student t and Plackett in the pair DE).

6. ESTIMATION OF JOINT EXTREME DEFAULTS: COMPARISON BETWEEN TRADITIONAL METHODS AND COPULAS

We now compare estimations of joint losses (default rates) assuming normal distributions (both univariate and multivariate) to evaluations based on Copula Theory. It is expected that, in general, approaches based on copulas give more efficient assessment of joint extremely-high losses. To test this hypothesis, we calculated the probability of default rates, D_A and D_B in segments A and B respectively, being simultaneously above specific levels (values) "a" and "b" as follows:

- Assuming normality: $\Pr(D_A > a, D_B > b) = 1 - \Phi(a) - \Phi(b) + \Phi(a,b)$

where Φ indicates the *cdf* of a normal distribution; and

- Using the "best" estimated survival copula: $\Pr(D_A > a, D_B > b) = \hat{C}[1 - F_A(a), 1 - F_B(b)]$

where $\hat{C}$ is a survival copula (see Section 2.1), i.e. links "survival ranks": 1-$F(.)$; F_A and F_B are the *cdfs* of the (unknown) distributions of default rates D_A and D_B, respectively.

Given that the dataset has 24 observations, the following proportions of "extremely" high levels (percentiles) of default rates were selected: 4.17% (1/24), 8.33% (2/24), 12.50% (3/24), 16.67% (4/24), 20.83% (5/24), and 25% (6/24). Thus, for each pair of segments we compare estimations of potential joint losses in those highest levels. For instance, the likelihood that the 4.17% highest default rates in segment A happen at the same time that the 4.17% highest default rates in segment B, and so on.

As the best-fit copulas were estimated according to two approaches (based on the complete default distributions and in their right tails), the survival copulas used to evaluate the joint occurrences were also determined following both strategies.

6.1 Survival copulas estimated considering complete distributions of default rates

The results pertaining to the survival copulas estimated from the whole default distributions are shown in Table 6 where the first column exhibits the *proportion* of the highest defaults (not the default rates themselves). The columns labeled "Dataset" give the proportion of joint default rates observed in the credit card portfolio at the respective levels.

According to Table 6, in 63.33% of the scenarios, when compared to the results obtained from the assumption of normality, the approximations derived from copulas were closer to the ratio of simultaneous high default rates observed in the credit card portfolio. That is, in 38 out of the 60 situations represented in Table 6, the absolute difference between columns "Dataset" and "Copula" was smaller than the absolute difference between columns "Dataset" and "Normal". Alas, such difference was not statistically significant at the 5% level and was significant only for pairs AD and CD at the 10% level.

[Insert Table 6 here]

The copula approach resulted in higher underestimation rate (16.67%) than calculations assuming normality (11.67%). Hence this empirical analysis does not support the hypothesis that evaluations based on normality assumptions are prone to underestimate "extreme" joint defaults since their underestimation ratio was relatively low (11.67%) compared to the alternative method (16.67%)[14]. This is likely due to the short period covered by the dataset which virtually ruled out the probability of *joint* "extreme" occurrences (if we define "extreme" as, for example, above the 95th percentile, such "extreme" events would take place only if the highest default rate of each segment happened in the same month).

However if we compare, for each pair, the mean absolute difference between the columns "Normal" and "Dataset" in Table 6 to the mean absolute difference between the columns "Copula" and "Dataset", we conclude that the latter difference is smaller for nine of the ten pairs[15]. This is evidence that if we need to choose one of the approaches (normality- or copula-based) to estimate the likelihood of joint defaults at *all* the six levels tested, we are better off if we opt for the copula method because, in general, estimates from this method will be closer to the observed default rates than results derived from the bivariate normal distribution.

[14] This conclusion is also valid if we analyze more extreme points in the right tail of each default distribution. If we consider, for example, potential joint occurrences among the 91.67% highest default rates (equivalent to the two highest observations in each segment), 20% of the estimates based on copulas were below the observed joint occurrences whilst 15% of the calculations related to the bivariate normal distribution resulted in underestimations.

[15] To keep Table 6 as simple as possible, these absolute differences are not displayed there. In each pair of segments, the mean difference is naturally the summation of the differences at the six risk levels divided by six.

6.2 Survival copulas estimated considering the right tail of the default rate distributions

The comparison between copula and traditional methods was repeated by using survival copulas estimations based only on the upper tail (above the 75th percentile) of default rate distributions (method similar the one presented in Section 5.2). The results are displayed in Table 7. Other percentiles higher than the 75th percentile were tested as the cutoff to define the right tail (e.g. the 87th and the 92nd) but the analyses of simultaneous extreme defaults based on their best-fit copulas were not fruitful because more than half of the pairs did not have joint occurrences at those levels.

[Insert Table 7 here]

The copula estimations were closer to the real default rates (observed in the dataset) in 70% of the cases[16] but presented a higher underestimation rate (26.67%) than the normality-based estimations (11.67%)[17]. As in the analysis of the previous item (for survival copulas estimated from the whole default distributions), this finding does not corroborate the idea that evaluations from normality assumptions tend to underestimate the odds of extreme events. Again, this failure in confirming that hypothesis is likely due to the short range covered by the dataset which excludes the possibility of checking simultaneous occurrences in the very tail of the distributions (for instance, the 1% highest default rates).

Nevertheless, taking into consideration *each pair separately*, the average of the difference between estimations and observed default rates was smaller for the copula approach in all ten pairs. So, as in Section 6.1, if we have to select one model to estimate potential joint losses at several levels, we tend to obtain better overall approximations from the copula method than from the approach founded on the assumption of normality.

7. FINAL COMMENTS

Copula Theory has been employed in credit risk analyses but, to the best of our knowledge, this is the first investigation to present an empirical study of copulas for consumer loans. Moreover

[16] Notwithstanding the difference between the traditional and the copula-based estimations was not significant at the 5% level for none of the pairs and significant at the 10% level only for pairs AD, BD and CD.

[17] Tests limited to the two farthest points in the right tail of the distributions (91.67% highest default rates) confirmed this fact. The underestimation ratios were 20% and 15% for the copula-based and the normality-based approaches respectively.

we test five copulas that are not typically considered in the literature and three of them (Galambos, Hüsler-Reiss, and Plackett)[18] were found to be the most adequate to represent the dependence between some segments. Given that the main objective is to find dependence structures (copulas) that yield more precise estimates of simultaneous high losses, another innovation was the inclusion of goodness-of-fit (GoF) tests based exclusively on the right tails of the default distributions (along with the complete distributions, which is often done).

As for the usual strategy (using the whole default distributions), among the ten segments investigated, eight present tail dependence[19] (i.e. higher association across extreme occurrences than across moderate events), from which five have upper-tail dependence, indicating that higher losses are more correlated. This suggests that, especially in downturns, the Financial Institution considered here is subject to losses in the credit card portfolio higher than those assumed by traditional models. Only in one pair is the dependence expressed by the Gaussian copula yet that is implicit in many models used nowadays.

With regard to the alternative strategy (GoF based on the right-tail of default distributions), seven pairs have tail dependence; six of them are right-tail dependent and one is left-tail dependent[20]. This confirms the conclusion that most of the pairs tend to be more associated when default rates are higher (i.e. in unfavorable economic scenarios). None of the pairs is represented by the Gaussian copula.

Although the Gaussian copula (the basis of some traditional credit risk models) cannot be rejected for most of the pairs (since they cannot be statistically rejected due to high p-values), the most important conclusion of this study is that the dependence across credit card loans can be better expressed by other copula families that indicate stronger association across high default rates than by the dependence inferred from the Gaussian copula. Consequently, estimations of the probability of joint high default rates are likely to yield more realistic results when copulas other than the Gaussian are used.

The limitations of traditional credit models in terms of estimation of joint extreme losses refer not only to the assumption of univariate losses' distributions but also to the treatment of the *joint*

[18] These three families were found when the GoF tests were based on the complete distributions of the default rates. Four "atypical" copulas resulted from estimations supported by the right tail of the default distributions: Galambos, Hüsler-Reiss, Joe, and Plackett.

[19] Although the results for two of those pairs are not statistically significant.

[20] In these particular cases, the copulas express tail dependence in only one of the tails. So, the pairs stated as right-tail dependent do not present left-tail dependence and vice versa.

behavior across defaults. In the credit card portfolio analyzed, we show some examples of segments (A, B, and C) with distributions statistically close to normality whose dependence is far from the Gaussian copula implicit in traditional credit risk models.

The comparison between joint extreme losses estimations derived from normality assumptions and copulas followed those two GoF strategies mentioned above. We found a trade-off between these two approaches: the one based on the right-tail of default distributions selects copulas more representative of extreme defaults (which improves copula estimations of joint *high* defaults) at the expense of higher underestimation indices (which we want to avoid).

Nonetheless our conclusions are limited due to the short period covered by the dataset (24 months). Even though it includes some months with intense losses at the end of 2008 (the so-called "credit crunch"), which it is interesting to check a possible higher link among higher default rates, it does not have enough observations to generate potential joint losses in the extremely high tail of the distributions (98% or 99%, for example) where the biggest deficiency of traditional models seems to be. The limitation of our results becomes clear if we recall that most of the copulas tested could not be rejected and the difference between the estimations of joint extreme default rates based on copulas and the estimations based on the assumption of normality was not statistically significant (at the 5% significance level).

Therefore a natural extension of this work would be to apply the same procedure in a dataset covering a longer time horizon to verify the estimations of joint events at extreme levels and to obtain more significant results. In order to consolidate the use of copulas in consumer loans, the dependence structure and the probability of severe losses in other types of portfolios, e.g. mortgages and fixed term loans, should be assessed and compared to estimations from traditional models.

It is worth bearing in mind that the method used to estimate the copula parameters in this analysis assumes that the variables (default rates) do not present temporal dependence and their individual distributions are constant over the period analyzed. In future studies, techniques that take serial correlation and distribution changes into account should be employed.

The performance of the class of model selectors used here (goodness-of-fit tests for distributions) can be compared to the performance of other methods such as (Pseudo-) Likelihood Ratio tests and two information criteria tests (Akaike and Schwarz's Bayesian). This will be an important contribution to the comparison of copula selection methods given that, to our knowledge, the power studies concerning these methods have been restricted to goodness-of-fit tests (notably, Genest et al., 2009 and Berg, 2009).

Due to the diversity of copulas found to represent the association between pairs of segments, it is interesting to search for a combination of copulas that represents the heterogeneous dependence across all segments at once and is associated to the multivariate distribution of the whole portfolio. This topic can be addressed by employing Vine Copulas (which are combinations of bivariate copulas in order to assess higher-dimension dependence).

REFERENCES

Ang, A., G. Bekaert (2002). International Asset Allocation With Regime Shifts. *Review of Financial Studies*, 15, pp. 1137-1187.

Berg, Daniel (2009). Copula goodness-of-fit testing: an overview and power comparison. *Europe Journal of Finance*, Vol. 15, Issues 7 & 8, pp. 675-701.

Cherubini, U., E. Luciano, W. Vecchiato (2004). *Copula Methods in Finance*. West Sussex: John Wiley & Sons Inc.

Das, S., G. Geng (2006). Correlated Default Processes: A Criterion-Based Copula Approach. In: Fong, H. (ed.) *"The Credit Market Handbook. Advanced Modeling Issues"*. Hoboken, New Jersey: John Wiley & Sons, Inc., pp. 186-218.

Daul, S., E. De Giorgi, F. Lindskog, A. McNeil (2003). The grouped t-copula with an application to credit risk. *RISK*, 16, pp. 73-76.

Di Clemente, A., C. Romano (2004). Measuring and Optimizing Portfolio Credit Risk: A Copula-based Approach. *Economic Notes* (by Banca Monte dei Paschi di Siena SpA), Vol. 33, nº 3, pp. 325-357.

Durrleman, V., A. Nikeghbali, T. Roncalli (2000). Which copula is the right one? Available at SSRN: http://ssrn.com/abstract=1032545.

Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business*, 38(1), pp. 34-105.

Frey, R., A. McNeil (2001). Modelling Dependent Defaults. ETH, Eidgenössische Technische Hochschule Zürick, Department of Mathematics, doi: 10.3929/ethz-a-004504297.

Frey, R., A. McNeil, M. Nyfeler (2001). Copulas and Credit Models. *RISK*, October, pp. 111-114.

Genest C., A-C. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, July/August, pp. 347-368

Genest, C., B. Rémillard, D. Beaudoin (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44, pp. 199-213.

Hamerle, A., D. Rösch (2005). Misspecified Copulas in Credit Risk Models: How Good is Gaussian? *Journal of Risk*, 8 (1), pp. 41-58.

Hamilton, D., J. James, N. Webber (2001). Copulas methods and the Analysis of Credit Risk. Available at SSRN: http://ssrn.com/abstract=1014407.

Hull, J., A. White (2006). Valuing Credit Derivatives Using an Implied Copula Approach. *Journal of Derivatives*, Vol. 14, No. 2, pp. 8-28. (Previously titled: *The Perfect Copula*)

Jarque, C. M., A. K. Bera (1987). A test for normality of observations and regression residuals. *International Statistical Review*. Vol. 55, No. 2, pp. 163–172.

Joe, Harry (1997). *Multivariate Models and Dependence Concepts.* London, New York: Chapman & Hall/CRC.

Joe, H., J. Xu (1996). The estimation method of inference functions for margins for multivariate models. Technical Report, n. 166, Dept. of Statistics, University of British Columbia.

Kole, E., K. Koedijk, M. Verbeek (2007). Selecting copulas for risk management. *Journal of Banking and Finance*, 31, pp. 2405-2423.

Kostadinov, Krassimir (2005). Non-parametric Estimation of Elliptical Copulae With Application to Credit Risk. Working paper, Munich University Technology, http://www.ma.tum.de/stat.

Kupiec, Paul H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives*, Vol. 3, N^{o}2, pp. 73-84.

Li, David (2000). On Default Correlation: A Copula Function Approach. *Journal of Fixed Income,* 9, pp. 43-55.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4), pp. 394-419.

McNeil, A. J., R. Frey, P. Embrechts (2005). *Quantitative Risk Management. Concepts, Techniques and Tools.* Princeton, UK: Princeton University Press.

Melchiori, Mario R. (2003). Which Archimedean Copula is the right one? YieldCurve.com e-Journal, October. Available at SSRN: http://ssrn.com/abstract=1123135.

Nelsen, Roger B. (2006). *An Introduction to Copulas.* New York: Springer. 2nd edition.

Nikoloulopoulos, A., J. Joe, H. Li (2011). Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics & Data Analysis* (forthcoming).

Ning, Cathy (2010). Dependence Structure between the Equity Market and the Foreign Exchange Market – a Copula Approach. *Journal of International Money and Finance*, Vol. 29(5), pp. 743-759.

Patton, Andrew J. (2006). Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*. Vol. 47, nº 2, pp. 527-556.

Patton, Andrew J. (2009). Copula-Based Models for Financial Times Series. In: Mikosch, T., J-P. Kreiß, R. Davis and T. Andersen (eds.) "*Handbook of Financial Times Series*", Berlin: Springer, pp. 767-785.

Rosenberg, J. V., T. Schuermann (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics*, 79, pp. 569-614.

Schmidt, Rafael (2003). Credit risk modelling and estimation via elliptical copulae. In: Bohl, G., G. Nakhaeizadeh, S. T. Rachev, T. Ridder, K. H. Vollmer (eds.), "*Credit Risk - Measurement, Evaluation and Management*", Physica-Verlag Heidelberg, pp. 267-289.

Schönbucher, P., D. Schubert (2001). Copula-Dependent Default Risk in Intensity Models. Working paper, Department of Statistics, Bonn University.

Trivedi, P. K., D. M. Zimmer (2007). Copula Modeling: An Introduction for Practitioners. *Foundations and Trends® in Econometrics*, Vol. 1, No 1, pp. 1-111.

Weiß, Gregor N. F. (2009). Are Copula-GoF-Test of any Practical Use? Results from a Simulation Study. 22nd Australasian Finance and Banking Conference 2009. Available at SSRN: http://ssrn.com/abstract=1459654.

APPENDIX A

Best-fit copulas' parameters

(copulas estimated according to the complete default distributions)

TABLE A.1 Copula parameters estimated for pairs AB, AC, AD and AE (best-fit based on complete default distributions)

GoF APPROACH	AB		AC		AD		AE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Clayton	2.9580448	Clayton	3.0354896	Galambos	0.0125000	Student t	-0.4430353
Kendall's Transform	Plackett	7.6454102	Plackett	8.2162109	Galambos	0.0125000	Frank	-3.3524919
Rosenblatt's Transform	Frank	4.7446299	Clayton	3.0354896	Gumbel	1.0000014	Plackett	0.2063477

TABLE A.2 Copula parameters estimated for pairs BC, BD, and BE (best-fit based on complete default distributions)

GoF APPROACH	BC		BD		BE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Clayton	11.3340991	Hüsler-Reiss	0.6412109	Student t	-0.5058982
Kendall's Transform	Clayton	11.3340991	Galambos	0.0125000	Plackett	0.1311523
Rosenblatt's Transform	Plackett	161.3443359	Joe	1.1007813	Plackett	0.1311523

TABLE A.3 Copula parameters estimated for pairs CD, CE, and DE (best-fit based on complete default distributions)

GoF APPROACH	CD		CE		DE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Hüsler-Reiss	0.6699219	Gaussian	-0.4534993	Plackett	3.5805664
Kendall's Transform	Joe	1.1268555	Frank	-4.5301524	FGM	0.7535625
Rosenblatt's Transform	Joe	1.1268555	Plackett	0.0832031	Plackett	3.5805664

APPENDIX B

Best-fit copulas' parameters

(copulas estimated according to right tails of default distributions)

TABLE B.1 Copula parameters estimated for pairs AB, AC, AD and AE
(best-fit based on right tails of default distributions)

GoF APPROACH	AB		AC		AD		AE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Galambos	0.0125000	Joe	1.2684570	Galambos	0.0125000	Frank	-3.3524919
Kendall's Transform	Galambos	0.0125000	Galambos	0.0125000	Gumbel	1.0000014	Clayton	0.0000015
Rosenblatt's Transform	Frank	4.7446299	Clayton	3.0354896	Gumbel	1.0000014	Plackett	0.2063477

TABLE B.2 Copula parameters estimated for pairs BC, BD, and BE
(best-fit based on right tails of default distributions)

GoF APPROACH	BC		BD		BE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Clayton	11.3340991	Gumbel	1.1175104	Frank	-3.9901586
Kendall's Transform	Student t	0.9622511	Clayton	0.7649150	Clayton	0.0000015
Rosenblatt's Transform	Plackett	161.3443359	Joe	1.1007813	Plackett	0.1311523

TABLE B.3 Copula parameters estimated for pairs CD, CE, and DE
(best-fit based on right tails of default distributions)

GoF APPROACH	CD		CE		DE	
	Copula	Parameter	Copula	Parameter	Copula	Parameter
Empirical Copula	Joe	1.1268555	Plackett	0.0832031	Hüsler-Reiss	0.0999999
Kendall's Transform	Clayton	0.8203320	Clayton	0.0000015	Hüsler-Reiss	0.0999999
Rosenblatt's Transform	Joe	1.1268555	Plackett	0.0832031	Plackett	3.5805664

Table 1: Summary statistics of default rates for the five segments of the credit card portfolio

SEGMENT	MEAN	STD DEVIATION	SKEWNESS	KURTOSIS
A	0.00015	0.00009	-0.22504	2.55931
B	0.00056	0.00035	-0.04654	2.17929
C	0.00295	0.00141	-0.57593	2.75557
D	0.01117	0.00270	-2.78037	12.64415
E	0.03226	0.01873	2.12537	8.45527

Data refers to April/2007 – March/2009.

Table 2: Jarque-Bera test for the default rates' segments

SEGMENT	CAN VALUES BE APPROXIMATED TO NORMAL?	JARQUE-BERA STATISTICS
A	Yes	0.39678
B	Yes	0.68222
C	Yes	1.38653
D	No	123.93139
E	No	47.82875

This test checks if the normal distribution is a good approximation for the default rates.

Table 3: Candidate copulas and their respective features

COPULA	DEPENDENCE STRUCTURE
Gaussian	Symmetric dependence without tail dependence
Frank	Symmetric dependence without tail dependence
FGM(*)	Symmetric dependence without tail dependence
Placket	Symmetric dependence without tail dependence
Student t	Symmetric dependence with (lower and upper) tail dependence
Clayton	Left (lower) tail dependence
Gumbel	Right (upper) tail dependence
Galambos	Right (upper) tail dependence
Hüsler-Reiss	Right (upper) tail dependence
Joe	Right (upper) tail dependence

(*) FGM stands for Farlie-Gumbel-Morgenstern.

Table 4: Best-fit copulas based on the complete distributions

SEGMENTS	A	B	C	D	E
A	1	Clayton** (Plackett**) [Frank]	Clayton* (Plackett*) [Clayton]	Galambos (Galambos) [Gumbel]	Student t (Frank) [Plackett]
B	0.7375 (0.4652)	1	Clayton (Clayton) [Plackett]	Hüsler-Reiss (Galambos) [Joe]	Student t (Plackett) [Plackett]
C	0.7888 (0.5146)	0.9536 (0.8577)	1	Hüsler-Reiss (Joe) [Joe]	Gaussian (Frank) [Plackett]
D	0.3730 (0.0036)	0.4598 (0.1421)	0.5653 (0.1851)	1	Plackett (FGM) [Plackett]
E	-0.4966 (-0.3679)	-0.4916 (-0.4335)	-0.5217 (-0.4537)	0.1241 (0.2464)	1

Best-fit copulas (upper-right triangle) and dependence measures (lower-left triangle) for default rates of pairs of segments (estimation based on the best-fit of complete distributions). The dependence measures are the linear correlation (above) and the Kendall's tau (in parenthesis). The copulas displayed for each pair of segments are respectively based on Empirical Copula, Kendall's Transform (in parenthesis), and Rosenblatt's Transform (in square brackets).

** and * indicate that the "best" copula found can be rejected at the 5% and 10% levels, respectively.

Table 5: Best-fit copulas based on best fit to right-hand tails

SEGMENTS	A	B	C	D	E
A	1	Galambos (Galambos) [Frank]	Joe (Galambos) [Clayton]	Galambos (Gumbel) [Gumbel]	Frank (Clayton) [Plackett]
B	0.7375 (0.4652)	1	Clayton (Student t) [Plackett]	Gumbel (Clayton) [Joe]	Frank (Clayton) [Plackett]
C	0.7888 (0.5146)	0.9536 (0.8577)	1	Joe (Clayton) [Joe]	Plackett (Clayton) [Plackett]
D	0.3730 (0.0036)	0.4598 (0.1421)	0.5653 (0.1851)	1	Hüsler-Reiss (Hüsler-Reiss) [Plackett]
E	-0.4966 (-0.3679)	-0.4916 (-0.4335)	-0.5217 (-0.4537)	0.1241 (0.2464)	1

Best-fit copulas (upper-right triangle) and dependence measures (lower-left triangle) for default rates of pairs of segments (correlation for complete default distributions and copula estimation based on the best fit of right tails). The dependence measures are the linear correlation (above) and the Kendall's tau (in parenthesis). The copulas displayed for each pair of segments are respectively based on Empirical Copula, Kendall's Transform (in parenthesis), and Rosenblatt's Transform (in square brackets).

Table 6: Comparisons of predicted joint extreme default rates using entire samples

Panel A: Pairs AB and AC

Proportion of highest losses	AB (Joe)			AC (Gumbel)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.00000	0.00996	0.00579	0.00000	0.01429	0.01219
8.33%	0.00000	0.03390	0.02110	0.00000	0.04202	0.03188
12.50%	0.04167	0.05466	0.04363	0.04167	0.06583	0.05593
16.67%	0.04167	0.08998	0.07180	0.04167	0.10950	0.08335
20.83%	0.04167	0.09485	0.10445	0.08333	0.12797	0.11358
25.00%	0.12500	0.11761	0.14070	0.12500	0.15524	0.14625

Panel B: Pairs AD and AE

Proportion of highest losses	AD (Galambos)			AE (t)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.00000	0.00513	0.00174	0.00000	0.00000	0.00074
8.33%	0.00000	0.03551	0.00694	0.00000	0.00006	0.00247
12.50%	0.00000	0.04552	0.01563	0.00000	0.00307	0.00547
16.67%	0.00000	0.09269	0.02778	0.00000	0.01481	0.01017
20.83%	0.04167	0.10413	0.04340	0.00000	0.02134	0.01706
25.00%	0.04167	0.14683	0.06250	0.00000	0.05369	0.02672

Panel C: Pairs BC and BD

Proportion of highest losses	BC (Joe)			BD (t)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.04167	0.02615	0.01446	0.00000	0.00783	0.00720
8.33%	0.08333	0.05749	0.04447	0.04167	0.03707	0.01838
12.50%	0.08333	0.10913	0.08128	0.04167	0.06548	0.03276
16.67%	0.08333	0.11965	0.12152	0.04167	0.07776	0.05012
20.83%	0.16667	0.12932	0.16359	0.04167	0.08322	0.07039
25.00%	0.16667	0.14136	0.20665	0.04167	0.09841	0.09352

Panel D: Pairs BE and CD

Proportion of highest losses	BE (t)			CD (Clayton)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.00000	0.00000	0.00113	0.00000	0.01400	0.00599
8.33%	0.00000	0.00005	0.00307	0.04167	0.05045	0.01583
12.50%	0.00000	0.00503	0.00601	0.04167	0.08756	0.02871
16.67%	0.00000	0.00873	0.01026	0.04167	0.10206	0.04444
20.83%	0.00000	0.01177	0.01627	0.04167	0.12320	0.06296
25.00%	0.00000	0.02265	0.02459	0.08333	0.13932	0.08423

Panel E: Pairs CE and DE

Proportion of highest losses	CE (Gaussian)			DE (Plackett)		
	Dataset	**Normal**	**Copula**	**Dataset**	**Normal**	**Copula**
4.17%	0.00000	0.00000	0.00006	0.00000	0.00002	0.00517
8.33%	0.00000	0.00005	0.00060	0.00000	0.00923	0.01797
12.50%	0.00000	0.00577	0.00231	0.00000	0.06704	0.03605
16.67%	0.00000	0.01001	0.00590	0.00000	0.09455	0.05816
20.83%	0.00000	0.01761	0.01209	0.04167	0.11596	0.08357
25.00%	0.04167	0.03167	0.02155	0.12500	0.16891	0.11179

Comparison between estimations of likelihood of joint extreme high default rates (normality vs. copulas). The survival copulas are informed in parenthesis after the names of the pairs and were estimated based on entire distributions of default rates.

Table 7: Comparisons of predicted joint extreme default rates using tail distributions

Panel A: Pairs AB and AC

Proportion of highest losses	AB (Clayton)			AC (Clayton)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.00000	0.00996	0.01015	0.00000	0.01429	0.01101
8.33%	0.00000	0.03390	0.02350	0.00000	0.04202	0.02504
12.50%	0.04167	0.05466	0.03934	0.04167	0.06583	0.04145
16.67%	0.04167	0.08998	0.05753	0.04167	0.10950	0.06010
20.83%	0.04167	0.09485	0.07800	0.08333	0.12797	0.08095
25.00%	0.12500	0.11761	0.10076	0.12500	0.15524	0.10399

Panel B: Pairs AD and AE

Proportion of highest losses	AD (Galambos)			AE (Frank)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.00000	0.00513	0.00174	0.00000	0.00000	0.00024
8.33%	0.00000	0.03551	0.00694	0.00000	0.00006	0.00112
12.50%	0.00000	0.04552	0.01563	0.00000	0.00307	0.00292
16.67%	0.00000	0.09269	0.02778	0.00000	0.01481	0.00600
20.83%	0.04167	0.10413	0.04340	0.00000	0.02134	0.01085
25.00%	0.04167	0.14683	0.06250	0.00000	0.05369	0.01806

Panel C: Pairs BC and BD

Proportion of highest losses	BC (Joe)			BD (Galambos)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.04167	0.02615	0.01446	0.00000	0.00783	0.00174
8.33%	0.08333	0.05749	0.04447	0.04167	0.03707	0.00694
12.50%	0.08333	0.10913	0.08128	0.04167	0.06548	0.01563
16.67%	0.08333	0.11965	0.12152	0.04167	0.07776	0.02778
20.83%	0.16667	0.12932	0.16359	0.04167	0.08322	0.04340
25.00%	0.16667	0.14136	0.20665	0.04167	0.09841	0.06250

Panel D: Pairs BE and CD

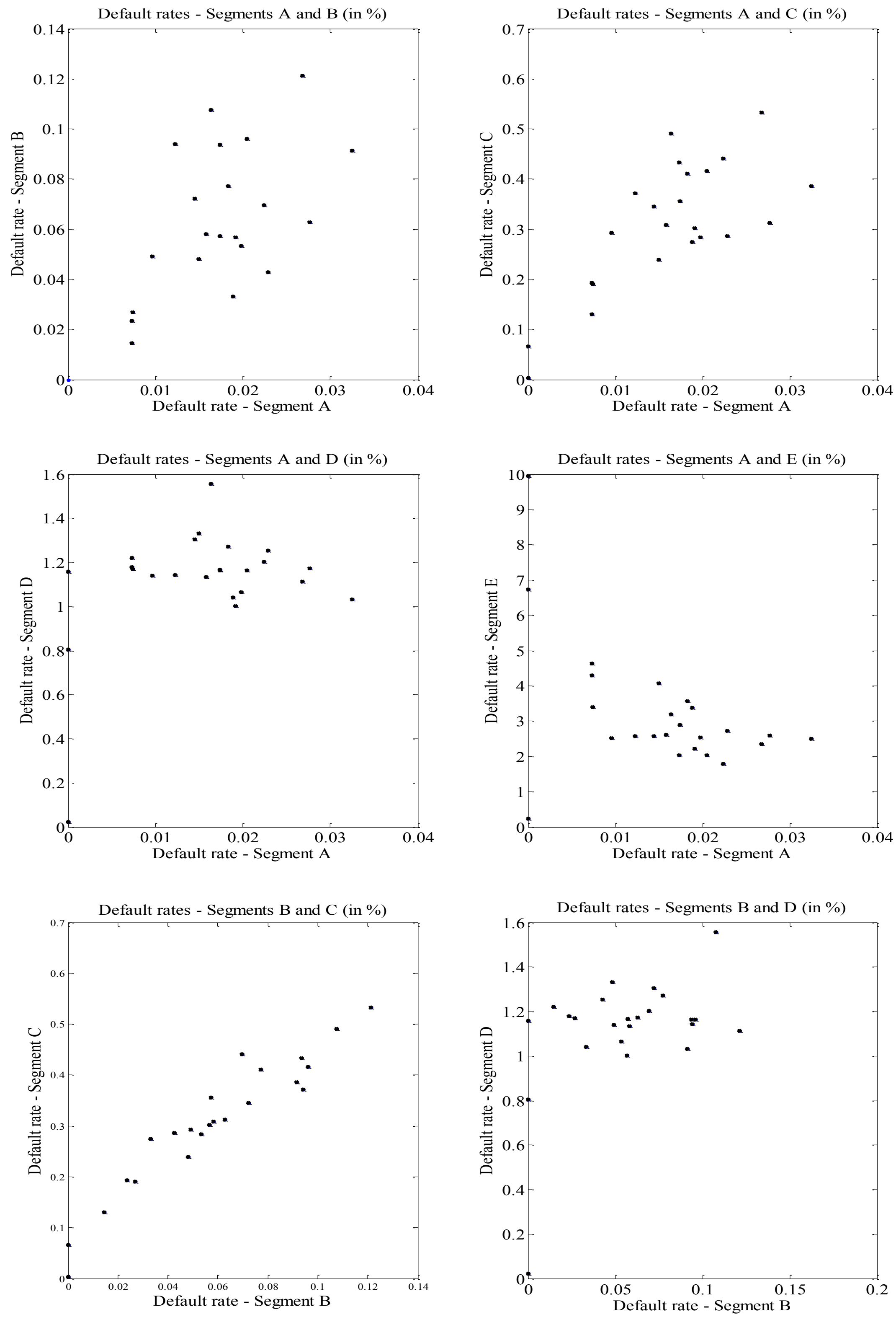
Proportion of highest losses	BE (Frank)			CD (Clayton)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.00000	0.00000	0.00015	0.00000	0.01400	0.00599
8.33%	0.00000	0.00005	0.00073	0.04167	0.05045	0.01583
12.50%	0.00000	0.00503	0.00197	0.04167	0.08756	0.02871
16.67%	0.00000	0.00873	0.00418	0.04167	0.10206	0.04444
20.83%	0.00000	0.01177	0.00781	0.04167	0.12320	0.06296
25.00%	0.00000	0.02265	0.01347	0.08333	0.13932	0.08423

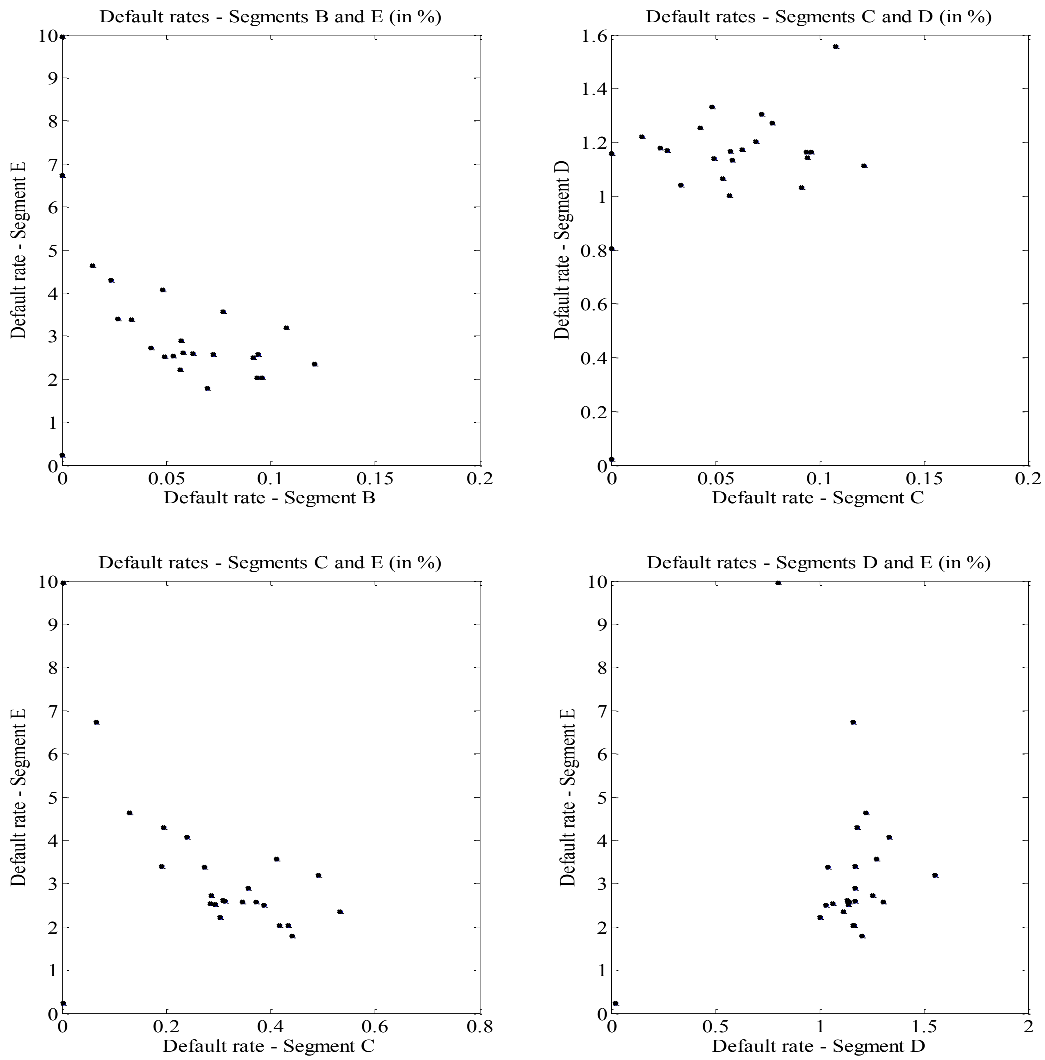
Panel E: Pairs CE and DE

Proportion of highest losses	CE (Gaussian)			DE (Joe)		
	Dataset	Normal	Copula	Dataset	Normal	Copula
4.17%	0.00000	0.00000	0.00006	0.00000	0.00002	0.00297
8.33%	0.00000	0.00005	0.00060	0.00000	0.00923	0.01153
12.50%	0.00000	0.00577	0.00231	0.00000	0.06704	0.02521
16.67%	0.00000	0.01001	0.00590	0.00000	0.09455	0.04361
20.83%	0.00000	0.01761	0.01209	0.04167	0.11596	0.06636
25.00%	0.04167	0.03167	0.02155	0.12500	0.16891	0.09314

Comparison between estimations of likelihood of joint extreme high default rates (normality vs. copulas). The survival copulas are informed in parenthesis after the names of the pairs and were estimated based on the right tails (above the 75th percentile) of the distributions of default rates.

Figure 1: Pairwise representation of the default rates in the credit segments analyzed




Default rates - Segments B and E (in %)
Default rate - Segment E
Default rate - Segment B
Default rates - Segments C and D (in %)
Default rate - Segment D
Default rate - Segment C
Default rates - Segments C and E (in %)
Default rate - Segment E
Default rate - Segment C
Default rates - Segments D and E (in %)
Default rate - Segment E
Default rate - Segment D